

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau



B3

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> :	A1	(11) International Publication Number: <b>WO 98/25947</b> (43) International Publication Date: 18 June 1998 (18.06.98)
C07H 21/04, C12Q 1/68, C12N 1/21, 15/70, G01N 33/53		
(21) International Application Number: PCT/US97/22703		(81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, UZ, VN, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).
(22) International Filing Date: 10 December 1997 (10.12.97)		
(30) Priority Data: 60/032,821 11 December 1996 (11.12.96) US		
(71) Applicant: BRISTOL-MYERS SQUIBB COMPANY [US/US]; P.O. Box 4000, Princeton, NJ 08543-4000 (US).		
(72) Inventor: KORNACKER, Michael, G.; 110 Nassau Street, Princeton, NJ 08540 (US).		
(74) Agents: GAUL, Timothy, J. et al.; Bristol-Myers Squibb Company, P.O. Box 4000, Princeton, NJ 08543-4000 (US).		

(54) Title: PROKARYOTIC TWO-HYBRID SYSTEM

(57) Abstract

A two-hybrid system that can detect homo- and heterodimeric protein interactions in *E. coli* and other cells. This system is useful for the same applications as a yeast two-hybrid system, i.e. interaction cloning, mapping protein interaction domains, analysing protein interactions, detecting protein interactions and detecting modulators thereof. The invention concerns a prokaryotic host cell comprising: a) a fusion protein having (i) a first DNA-binding domain and (ii) a first interacting domain; b) a fusion protein having (i) a second DNA-binding domain and (ii) a second interacting domain capable of binding to the first interacting domain; and c) a nucleic acid molecule having a reporter gene operatively linked to (i) a promoter, (ii) a first operator site capable of binding to the first DNA-binding domain, located upstream of the promoter, and (iii) a second operator site capable of binding the second DNA-binding domain, located downstream of the promoter of the reporter gene; wherein binding of the first interacting domain to the second interacting domain is signaled by altered expression of the reporter gene.

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## PROKARYOTIC TWO-HYBRID SYSTEM

### Field of the Invention

This invention relates to assays for protein interactions, fusion  
5 proteins, and host cells modified to comprise such proteins.

### Background of the Invention

Homo- and heterodimeric protein complexes mediate many cellular processes and abnormal protein interactions underly various medical  
10 conditions. Yan *et al.* (1995) *Cancer-Res.* 55: 3569-75. Research on such complexes has led to efforts to understand disease at the molecular level and to a search for small molecule effectors of such complexes. Such  
15 effectors could modulate protein interactions and are potential therapeutic agents. Gibbs & Oliff (1994) *Cell* 79: 193-198. Most often, such effectors have been identified using various biochemical and immunological *in vitro*  
approaches. The advantages of genetic approaches in drug discovery,  
however, have recently received increased attention. Liuzzi *et al.* (1994),  
*Nature* 372: 695-8. These advantages include both cost-effectiveness and simplicity. Only one such genetic system, the yeast-two hybrid system,  
20 currently meets all these criteria and is also equally suitable for the detection of both homo- and heterodimeric protein interactions. Another unique feature of the yeast two-hybrid system is its ability to detect the desired protein-protein interaction without interference by competing interactions. Fields & Song (1989) *Nature* 340: 245-6. The system has been  
25 successfully used for the analysis of protein interactions and for the isolation of interacting proteins through interaction cloning. For a review, see Allen *et al.* (1995), *Trends in Biochem. Sci.* 20: 511-16.

Although the yeast two-hybrid system has proven highly useful, it suffers from a number of limitations. Yeast is impermeable to many small  
30 molecules, which effectively prevents their evaluation in a yeast system. Higgins (1993) *Curr. Opin. in Cell Biol.* 5: 684-687. The yeast system also

requires nuclear localization of interacting proteins, which may lead to other complications.

- These problems can potentially be overcome with an E. coli two-hybrid system. E. coli strains can be hyperpermeable. Nakamura & Suganuma (1972) J. Bacteriol. 110: 329-35. One can use this hyperpermeability to maximize the number of small molecules that can be evaluated. In addition, E. coli has a rapid growth rate, permitting shorter turnaround times during drug screening. Furthermore, one can transform E. coli at high frequencies, facilitating interaction cloning.
- To date, only one E. coli system seems to have properties similar to the yeast two-hybrid system, but this system has only been shown to detect homodimerization of an E. coli protein and there is no published evidence that the system is sufficiently robust to be useful for major two-hybrid applications such as interaction cloning. Dove et al. (1997), Nature 386: 627-30. Other E. coli systems for detecting protein interactions genetically have also been reported. Doerr et al. (1991) Biochem. 30: 9657-64; Marchetti et al. (1995) J. Mol. Biol. 248: 541-50; Jappelli & Brenner (1996) J. Mol. Biol. 259: 575-8. Unlike the yeast two-hybrid system, however, these systems require homodimerization of at least one of the interacting proteins and detection of the desired protein-protein interaction can be subject to interference by competing other interactions. The art would benefit from an E. coli two-hybrid system that can detect a variety of protein interactions and which is sufficiently robust for interaction cloning.

25

#### Summary of the Invention

- The present invention relates to a two-hybrid system that can detect homo- and heterodimeric protein interactions in E. coli and other prokaryotic cells. This system is useful for the same applications as the yeast two-hybrid system; e.g., interaction cloning, in detecting and analyzing protein interactions and detecting modulators thereof. The invention concerns a prokaryotic host cell comprising:

- (a) a fusion protein having
  - (i) a first DNA-binding domain and
  - (ii) a first protein-interacting domain;
- 5 (b) a fusion protein having
  - (i) a second DNA-binding domain and
  - (ii) a second protein-interacting domain capable of binding to the first interacting domain; and
- (c) a nucleic acid molecule having a reporter gene operatively linked to
  - (i) a promoter,
  - (ii) a first operator site capable of binding to the first DNA-binding domain, located upstream of the promoter, and
  - (iii) a second operator site capable of binding the second DNA-binding domain, located downstream of the promoter and the first operator site.

Binding of the first interacting domain to the second interacting domain is signaled by altered (i.e., downregulated) expression of the reporter gene. In 20 the view of the present inventor, the fusion proteins form a loop together with the length of DNA between the binding sites. This loop apparently interferes with the activity of the RNA polymerase, perhaps by preventing its binding to the DNA.

The invention also concerns the nucleic acid molecule described in 25 subparagraph (c) above and a vector comprising this nucleic acid molecule.

The invention further concerns a process for detecting inhibition or enhancement of binding of a first interacting domain with a second interacting domain, which comprises (a) treating a culture of cells of Claim 1 with a test substance, and (b) screening for increased or decreased 30 expression of the reporter gene, respectively.

The invention further concerns an interaction cloning method—i.e., a process for detecting a cell comprising a test domain that interacts with a known domain, which comprises:

- (1) generating a library of cells in which each cell comprises:
  - 5 (a) a fusion protein having
    - (i) a first DNA-binding domain and
    - (ii) a known domain;
  - (b) a fusion protein having
    - (i) a second DNA-binding domain and
    - (ii) a test domain; and
  - 10 (c) a nucleic acid molecule having a reporter gene operatively linked to
    - (i) a promoter,
    - (ii) a first operator site capable of binding to the first
  - 15 (iii) a second operator site capable of binding the second DNA-binding domain, located downstream of the promoter, and
  - 20 (iv) a second operator site capable of binding the second DNA-binding domain, located downstream of the promoter and the first operator site;
- (2) detecting cells exhibiting altered expression of the reporter gene, which signals interaction between the known domain and the test domain.

The invention also concerns a number of preferred variations to the above-described host cell, nucleic acid, vector, and process. The preferred first DNA-binding domain is from AraC, and the preferred first operator site is an AraC operator. The preferred second DNA-binding protein is from LexA, and the preferred second operating site is a LexA half-operator site. In another variation, the nucleic acid comprises more than one LexA operator half site. In one further preferred variation, the nucleic acid comprises a binding site for a factor such as IHF between the operator sites

that causes the nucleic acid to bend upon binding. Further preferences appear in the detailed description that follows.

Although described below with *E. coli*, the invention encompasses a number of cell types. The invention may be adapted to other prokaryotic  
5 cells.

#### Brief Description of the Drawings

Figure 1: Diagrams of gene activation. (A) Yeast two-hybrid system.  
(B) Proposed mechanism for a two hybrid system using hybrid operators  
10 specific for the phage P22 and 434 repressor proteins. See specification  
hereinafter for details. X and Y represent interacting domains used to  
generate hybrid proteins; Activ: Activation domain; DNA: DNA binding  
domain; op: operator; RNAP: RNA polymerase.

Figure 2: Salient features of reporter plasmids and hybrid proteins.  
15 (A) Nucleotide sequence of the regulatory region upstream of the *lacZ* reporter gene in plasmid P3 (Reeder & Schleif (1993) *J. Mol. Biol.* 231: 205-218). Plasmid P3 contains an I1-I1 AraC operator and was used to construct the reporter plasmids in this study. The restriction sites for insertion of LexAOp and IHFOp are shown. -35 and -10 designate the AraC-dependent  
20 promoter that overlaps the promoter-proximal I1 site. +1 indicates the transcriptional start. See Reeder & Schleif (1993) *J. Mol. Biol.* 231: 205-218 for nucleotide sequence between *HindIII* and *lacZ*. Top strands of oligonucleotide LexAOp and IHFOp inserts are shown. Bold sequences indicate consensus LexA half sites. (B) Linear structures of hybrid proteins.  
25 Structures of wild-type AraC and LexA are shown on top. Numbers indicate boundary amino acid residues of original proteins used for hybrid construction. Dim: Dimerization domain. 'DNA' is defined in Figure 1.

Figure 3: Diagram outlining the mechanism of this *E. coli* two-hybrid system. See text for details.

30 Figure 4: Original assay results. (A-E) with reporter plasmid P3IHFLexAOpX; (F) with reporter plasmid P3IHFLexAOpmH. a: reporter strain alone; b: reporter strain with AraC construct; c: reporter strain with

AraC and LexA constructs. Hybrid pairs are indicated and repression ratios are given in brackets. AraC and LexA refer to the DNA binding domains of these proteins, AraC(wt) is wild-type AraC, mFos is Fos with the IVM54 mutation. Four colonies of each strain were assayed (Materials & Methods) 5 and average activities were used to calculate repression ratios (see text). Where tested, repression ratios did not significantly change when assaying more than four colonies. Vertical axis indicate LacZ activities as Miller units.

Figure 5: Dependence of repression on intracellular levels of LexA  
10 hybrids. (A) Jun-AraC/Fos-LexA. (B) Non-interacting control hybrid pair AraC/Fos-LexA. Aliquots of a diluted preculture of each strain were grown with IPTG and anhydrotetracycline (AT) to induce expression of AraC and LexA constructs, respectively. IPTG was used at 1mM. The amount of LexA hybrid was titrated with the following amounts of AT (ng/ml): a: 160, b:  
15 c: 80, d: 40, e: 0. Maximum repression ratios are indicated.

Figure 6: Diagram of disruption of dimeric AraC hybrids through heterodimerization with LexA hybrids. See text for details.

#### Detailed Description of the Invention

20 **Definition of terms**

The following definitions apply to the terms as used throughout this specification, unless otherwise limited in specific instances.

The term "prokaryotic host cell" refers to such genera and species as:

Escherichia coli

25 Salmonella

Klebsiella

Pseudomonas

Caulobacter

Rhizobium

30 and the like.

The term "first DNA-binding domain" refers to a polypeptide sequence capable of binding to an operator that could be inserted in the

region upstream from the reporter gene. Such domains can retain most or all of the protein from which they are derived or merely that portion that directly participates in DNA binding. Exemplary first DNA-binding domains may be derived from AraC, its homologues such as RhaS, XylS or 5 any other DNA binding proteins. Such proteins are described in Ramos *et al.* (1990) *Nucleic Acids Res.* 18: 2149.

The term "first operator site" refers to a nucleotide sequence capable of binding to the associated first DNA-binding domain. Such sequences are known in the art and can be inserted through recombinant DNA 10 techniques.

The term "second DNA-binding domain" refers to a polypeptide sequence capable of binding to an operator that could be inserted in the region downstream from the promoter. Such domains can retain most or all of the protein from which they are derived or merely that portion that 15 directly participates in DNA binding. Exemplary second DNA-binding domains may be derived from LexA, the phage lambda CI repressor, the repressors of phages P22 and 434, or any other DNA binding proteins that can bind DNA as fusion proteins.

The term "second operator site" refers to a nucleotide sequence 20 capable of binding to the associated second DNA-binding domain. Such sequences are known in the art and can be inserted through recombinant DNA techniques. Exemplary sequences are LexA operator sites, including half sites thereof. The second operator site is located between the promoter and translational start site for the reporter gene, with a location between 25 the transcriptional and translational start sites preferred.

The term "AraC binding domain" refers to a polypeptide deriveable from the AraC protein that is capable of binding to DNA. The AraC binding domain can be full-length AraC or any natural or modified fragment thereof that retains DNA binding activity. An exemplary AraC 30 binding domain has the sequence

ESLHPPMDNRVREACQYISDHLDNSFDIASVAQHVCLSPSR  
LSHLFRQQQLGISVLSWREDQRISQAKLLLSTTRMPIATVGRN

VGFDDQLYFSRVFKKCTGASPSEFRAGCEEKVNDVAVKLS  
(SEQ ID NO: 4).

The term "AraC operator site" refers to a nucleotide sequence capable of binding to AraC. Exemplary nucleotide sequences include

5 TAGCATTATCCATA  
(SEQ ID NO: 5).

The term "LexA binding domain" refers to a polypeptide deriveable from LexA that is capable of binding to DNA. The LexA binding domain can be full-length LexA or any natural or modified fragment thereof that 10 retains DNA binding activity. An exemplary LexA binding domain has the sequence

MKALTARQQEVFDLIRDHISQTGMPPTRAEIAQRLGFRSPNAA  
EEHLKALARKGVIEIVSGASRGIRLLQEEEGLPLVGRVAAGEPL  
(SEQ ID NO: 6).

15 The term "LexA operator site" refers to a nucleotide sequence capable of binding to LexA. LexA operators may be one or more consensus and/or nonconsensus half-sites, which may or may not be oriented as inverted repeats. Exemplary nucleotide sequences include LexA operator half sites, which can conform to the consensus sequence

20 TACTGTATAT  
(SEQ ID NO: 2).

Exemplary LexA operators include:

- LexA colE1 operator half site as triple direct repeats:

5' - AAAACCAGTGAAAACCAAGTGAACCAAGTG -3'

25 (SEQ ID NO: 8)

- LexA consensus half sites as quadruple direct repeats with 3 or 4 bp spacers shown in bold:

5' - ATATACAGTACCAATATACAGTACC  
ACATATACAGTACCAAATATACAGTA -3'

30 (SEQ ID NO: 9)

- Two full-length consensus LexA sites with 16 bp spacer shown in bold:

5' - TACTGTATATATACAGTACTTATAACG  
GCAAGTACTACTGTATATATACAGTA -3'

(SEQ ID NO: 10)

5       The term "reporter gene" refers to any gene whose expression provides a measurable signal. Exemplary reporter genes include the genes for  $\beta$ -galactosidase, antibiotic resistance genes (e.g., chloramphenicol transferase), and toxic genes (e.g., GATA-1 DNA binding domains, colicin lysis genes, and the like). Various other reporter genes are well known by 10 those having ordinary skill in the art.

The phrase "factors capable of causing the nucleic acid molecule to bend between the first and second operator sites" refers to any of a number of proteins that bind DNA and cause deformation of the linear axis of the helix. Exemplary factors are IHF, TF1, HU, CAP or intrinsically bent DNA 15 sequences such as C(A)<sub>5-6</sub>T, and the like. Geidschek *et al.* (1990) *J. Struc. Biol.* 104: 84-90; Hodges-Garcia *et al.* (1989) *J. Biol. Chem.* 264: 14621-3; Wu & Cruthers (1984) *Nature* 308: 509-13; Kahn & Crothers (1992) *Proc. Natl. Acad. Sci. USA* 89: 634307; Goodman *et al.* (1989) *Nature* 341: 244-51; Goodman *et al.* (1992), *Proc. Natl. Acad. Sci.* 89:11910-14. The associated 20 binding site may be any nucleotide sequence capable of such binding (e.g., the consensus binding site for IHF, the nucleotide sequence of SEQ ID NO: 3)

TX<sub>6</sub>ATXAX<sub>2</sub>TX<sub>2</sub>AXTXAAATCAAXAAGTTAX<sub>6</sub>A

SEQ ID NO: 3.

25   Process of preparation

Gene constructs

The nucleic acids used in the present invention may be prepared by recombinant nucleic acid methods. See, for example, the recombinant DNA methods of Nelles *et al.* (1987), *J. Biol. Chem.* 262: 10855. Exemplary 30 strains comprising such constructs areas described in the figure legends.

The DNA sequences may be derived from a variety of sources, including genomic DNA, subgenomic DNA, cDNA, synthetic DNA, and

combinations thereof. Genomic and cDNA may be obtained in a number of ways. Cells coding for the desired sequence may be isolated, the genomic DNA fragmented (e.g., by treatment with one or more restriction endonucleases), and the resulting fragments cloned, identified with a probe complementary to the desired sequence, and screened for the presence of a sequence coding for the desired activity.

For cDNA, the cDNA may be cloned and the resulting clone screened with a probe for cDNA coding for the desired region. Upon isolation of the desired clone, the cDNA may be manipulated in substantially the same manner as the genomic DNA.

To express the DNA sequences, transcriptional and translational signals recognized by an appropriate host are necessary. Alternatively, the promoter region from genomic DNA may be obtained in association with the DNA sequence for the fusion protein. To the extent that the host cells recognize the transcriptional regulatory and translational initiation signals associated with the fusion protein, the 5' region adjacent to the coding sequence may be retained and employed for transcriptional and translational regulation. This region typically will include those sequences involved with initiation of transcription and translation, such as the TATA box, capping sequence, CAAT sequence, and the like. Typically, this region will be at least about 150 base pairs long, more typically about 200 bp, and rarely exceeding about 1 to 2 kb.

The non-coding 3' region may be retained, as well, especially for its transcriptional termination regulatory sequences, such as the stop signal and polyadenylated region. In addition, the non-coding 3' region may also contain an enhancer. Where the transcriptional termination signals are not satisfactorily functional in the host cell, then a functional 3' region from a different gene may be substituted. In this method, the choice of the substituted 3' region would depend upon the cell system chosen for expression.

A wide variety of transcriptional and translational regulatory sequences may be employed, depending upon the nature of the host. The

transcriptional and translational regulatory sequences may be derived from viral sources (e.g., adenovirus, bovine papilloma virus, Simian virus, and the like) where the regulatory signals are derived from a gene that has a high level of expression in the host. Alternatively, promoters 5 from mammalian expression products (e.g., actin, collagen, myosin, and the like) may be employed. Transcriptional initiation regulatory signals may be selected that allow for repression or activation, so that expression of the genes can be modulated. One such controllable modulation technique is the use of regulatory signals that are temperature-sensitive, so 10 that expression can be repressed or initiated by changing the temperature. Another controllable modulation technique is the use of regulatory signals that are sensitive to certain chemicals.

To form the reporter or DNA binding domain-interacting domain chimeric gene constructs, DNA fragments may be ligated in accordance 15 with conventional techniques known in the art. Such techniques include use of restriction enzymes to convert sticky-ended fragments to blunt ends (or vice-versa), polymerases and nucleotides to fill in sticky ends to form blunt ends, alkaline phosphatase to avoid undesired ligations, and ligases to join fragments.

20 The construct for a DNA binding domain-interacting domain (e.g., for AraC or LexA and their respective fusion partners) may be joined together to form a single DNA segment or may be maintained as separate segments by themselves or in conjunction with vectors. The constructs may be introduced into a cell by transformation in conjunction with a 25 gene allowing for selection where the construct will become integrated into the host genome. Usually, the construct will be part of a vector having a replication system recognized by the host cell.

#### Expression vectors

30 Expression vehicles for production of the molecules of the invention include plasmids or other vectors. In general, such vectors contain control sequences that allow expression in various types of hosts, including prokaryotes. Suitable expression vectors containing the desired

coding and control sequences may be constructed using standard recombinant DNA techniques known in the art, many of which are described in Sambrook *et al.* (1989), Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor Laboratory, Cold Spring

5 Habor, NY.

An expression vector as contemplated by the present invention is at least capable of directing the replication of the reporter gene construct and the replication and expression of the DNA binding domain-interacting domain construct. One class of vectors utilizes DNA elements that provide 10 autonomously replicating extrachromosomal plasmids derived from animal viruses (e.g., bovine papilloma virus, polyomavirus, adenovirus, or SV40). A second class of vectors relies upon the integration of the desired gene sequences into the host cell chromosome.

Expression vectors useful in the present invention typically contain 15 an origin of replication, a promoter located 5' to (i.e., upstream of) the DNA sequence to be expressed, and a transcription termination sequence. Suitable origins of replication include, for example, the ColE1, pSC101, SV4O and M13 origins of replication. Suitable termination sequences include, for example, the bovine growth hormone, SV40, lacZ and 20 AcMNPV polyhedral polyadenylation signals. Suitable promoters include, for example, the cytomegalovirus promoter, the lacZ promoter, the gal 10 promoter and the AcMNPV polyhedral promoter. The promoter sequence may also be inducible, to allow modulation of expression (e.g., by the presence or absence of nutrients or other inducers in the growth medium). 25 One example is the lac operon obtained from bacteriophage lambda plac5, which can be induced by IPTG.

The expression vectors may also include other regulatory sequences for optimal expression of the desired product. Such sequences include 30 stability leader sequences, which provide for stability of the expression product; secretory leader sequences, which provide for secretion of the expression product; enhancers, which upregulate the expression of the DNA sequence; and restriction enzyme recognition sequences, which

provide sites for cleavage by restriction endonucleases. All of these materials are known in the art and are commercially available. See, for example, Okayama (1983), Mol. Cell. Biol., 3: 280.

- A suitable expression vector may also include marking sequences,
- 5 which allow phenotypic selection of transformed host cells. Such a marker may provide prototrophy to an auxotrophic host, biocide resistance (e.g., antibiotic resistance) and the like. The selectable marker gene can either be directly linked to the DNA gene sequences to be expressed, or introduced into the same cell by co-transfection. Examples of selectable markers
- 10 include neomycin, ampicillin, hygromycin resistance and the like.

The characteristics of the actual expression vector used must be compatible with the host cell that is to be employed. Suitable commercially available expression vectors into which the DNA sequences of the present invention may be inserted include pSPORT, pBluescriptIISK, the

15 baculovirus expression vector pBlueBac, and the prokaryotic expression vector pcDNAII, all of which may be obtained from Invitrogen Corp., San Diego, CA.

#### Host cells

The present invention additionally concerns hosts for the chimeric gene constructs and reporter plasmids. Suitable host cells include prokaryotic cells in which endogenous activities do not interfere with the function of the two-hybrid system. Such E. coli strains include MC1061, DH5a, Y1090 and JM101.

Expression vectors may be introduced into host cells by various methods known in the art. For example, transfection of host cells with expression vectors can be carried out by the calcium phosphate precipitation method. However, other methods for introducing expression vectors into host cells, for example, electroporation, liposomal fusion, nuclear injection, and viral or phage infection can also be employed. Host cells containing an expression vector may be identified by any of a number of methods known in the art.

### Detailed Description of Preferred Embodiments

Below are detailed descriptions of specific embodiments of the present invention. These embodiments are exemplary and serve to illustrate the broad applicability of the present invention.

5       The yeast two-hybrid system is a powerful genetic tool for detecting and analyzing homo- and heterodimeric protein interactions. Here we describe a mechanistically different *E. coli* two-hybrid system where protein interaction represses reporter gene activity. The system employs hybrids of two *E. coli* proteins, the activator protein AraC and the repressor 10 protein LexA. AraC hybrids alone activate a *lacZ* reporter gene from a high affinity operator directly upstream of the *araBAD* promoter that controls *lacZ*. LexA hybrids can repress this activation by up to 50-fold, but only if the hybrids have cognate interaction modules and the reporter plasmid contains LexA operator site(s) several helical turns downstream of the 15 promoter in the 5' untranslated leader of *lacZ*. These requirements suggest that repression occurs because DNA looping between operator-bound heterodimeric hybrids excludes RNA polymerase from the *lacZ* promoter inside the loop. Consistent with a mechanism based on DNA looping, repression is enhanced with the *E. coli* DNA bending protein IHF bound 20 between the operators. The results described here, suggest that the *E. coli* two-hybrid system will permit highly efficient interaction cloning and promises to be particularly useful for drug discovery and the analysis of interactions between bacterial proteins.

### MATERIALS AND METHODS

25       Standard cloning and immunoblotting techniques were used. Sambrook *et al.* (1989), *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.). Miller units of LacZ activity were calculated from a time course of 9 readings/sample on a BioTekTM EL-340 96-well microplate reader at 30 °C. 30 and with Chlorophenolred-b-D-galactopyranoside (CPRG; Boehringer Mannheim) as a substrate. Miller (1972) *Experiments in molecular genetics*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY; Menzel

(1989) Anal. Biochem. **181**, 40-50. For assays, E. coli MC1061 containing expression- and/or reporter plasmids was grown for at least 7 generations in LB-broth in microtiter plates in the presence of inducer (IPTG, Fisher Biotech; anhydrotetracycline, Acros) at 37 °C to A600=0.2-0.4. Meissner *et al.* (1987) Proc. Natl. Acad. Sci. USA **84**, 4171-4175. Under these conditions, LacZ activities are generally about 5-fold lower than reported previously. Bustos & Schleif (1993), Proc. Natl. Acad. Sci. USA **90**: 5638-42; Reeder, T. & Schleif, R. (1993) J. Mol. Biol. **231**, 205-218. HinD::cat was tested in strain TR321 after phage P1 transduction from strain HN1069. Bustos & Schleif (1993), Proc. Natl. Acad. Sci. USA **90**: 5638-42; Flamm , E. L. & Weisberg, R. A. (1985) J. Mol. Biol. **183**, 117-128.

Reporter plasmids are based on plasmid P3 (ColE1, Amp<sup>r</sup>) (Figure 2A; 11) or pMS421 (pSC101, Spec<sup>r</sup>). Churchward, G., Belin, D. & Nagamine, Y. (1984) Gene **31**, 165-171. LexA and IHF sites (Figure 2A) are oligonucleotide inserts where the mutant IHF site is in:

5'-CATAAACGAGCATTGCTGCGATATTGCAGCAAGGGAGCT-3'  
(SEQ ID NO: 7).

See Figure 2B for structures of hybrid proteins. araC constructs are under the IPTG-inducible pSE380 trp/lac promoter on p15A replicons (Cm<sup>r</sup> or Km<sup>r</sup>; Invitrogen. Bustos & Schleif (1993), Proc. Natl. Acad. Sci. USA **90**: 5638-42. LexA<sub>DNA</sub> constructs are under a second pSE380 regulatory element downstream and in tandem with the araC construct. Assays involving LexA<sub>DNA</sub> were in strains with pP3-based reporter plasmids, p15A-based hybrid constructs and pMS421 which carries lacI<sup>q</sup>. For assays with hybrids of full-length LexA, reporter constructs were on pMS421 and lexA constructs under the anhydrotetracycline-inducible tet promoter in pASK75 (ColE1, Amp<sup>r</sup>. Skerra, A. (1994) Gene **151**, 131-135. Zip-AraC<sub>DNA</sub>, zip-LexA<sub>DNA</sub>, AraC<sub>DNA</sub> and AraC are from pGB009, pGB004, pGB021b and pGB020, respectively. Bustos & Schleif (1993), Proc. Natl. Acad. Sci. USA **90**: 5638-42. Mouse c-fos and c-jun contain residues 126-218 and residues 208-324, respectively. Kouzarides, T. & Ziff, E. (1989) Nature **340**, 568-571. See Ryseck, R. P., Kovary, K. & Bravo, R. (1990) Oncogene **5**, 1091-1093 for Fos

mutation IVM54. E6 and E6AP domains are residues 8-158 and 351-451, respectively. Huibregtse, J. M., Scheffner, M. & Howley, P. (1991) EMBO J. 10, 4129-4135.

## 5 RESULTS

### Activation of a lac reporter gene by heterodimeric E. coli AraC and LexA hybrid proteins in an E. coli two-hybrid system.

Ptashne and coworkers described a general approach towards a bacterial two-hybrid system, using as models the repressor proteins of bacteriophages P22 and 434. Hollis et al. (1988), Proc. Natl. Acad. Sci. USA 85: 5834-8. These repressors only bind their operators as dimers and they can heterodimerize, but they do not recognize each others operators. However, heterodimers can act as repressors with hybrid operators where each half site is specific for one of the repressors (Figure 1B). These results suggested a two-hybrid system where repression from a hybrid operator would indicate interaction of other proteins when they replace the dimerization domains of the phage P22 and 434 repressors.

We chose the E. coli AraC and LexA proteins for a similar activation-approach in an E. coli two-hybrid system. AraC is a transcriptional activator that binds as a dimer to an operator of two tandem half sites, I1 and I2, just upstream of the AraC dependent araBAD promoter (Figure 1B). Schleif (1996) in Escherichia coli and Salmonella typhimurium. Cellular and Molecular Biology, (Neidhardt, ed.), ASM Press, Washington, D. C., 1300-9. LexA is a transcriptional repressor which cooperatively binds as a dimer to an operator of two inverted half sites. LexA has been used to provide the DNA binding function in the yeast two-hybrid system. Allen et al. (1995), Trends in Biochem. Sci. 20: 511-6. Both AraC and LexA have dimerization domains that can be functionally replaced by self-associating foreign proteins, such as the c/EBP leucine zipper (zip) to generate zip-AraCDNA and zip-LexADNA hybrids. Bustos & Schleif (1993), Proc. Natl. Acad. Sci. USA 90: 5638-42.

Several lines of indirect evidence suggest that the AraC-RNA polymerase contacts necessary for activation of araBAD may only require the promoter-proximal subunit of the dimer present at the low affinity I2 half site. Schleif (1996) in Escherichia coli and Salmonella typhimurium.

5    5    Cellular and Molecular Biology, (Neidhardt, ed.), ASM Press, Washington, D. C., 1300-9. I2 is accessed cooperatively when the other dimer subunit is bound at the promoter-distal high affinity I1 site. We reasoned that zip-AraCDNA/zip-LexADNA heterodimers might activate araBAD from a hybrid operator when a LexA half-site (LexAOp) replaces the AraC I1 site

10    10    (see Figure 2 for araBAD-lac reporter plasmid and structures of hybrids). Although each hybrid can homodimerize, the LexA hybrid homodimer, however, should be unable to bind to the hybrid LexAOp-I2 operator because its cooperative binding requires two LexA half sites. Bustos & Schleif (1993), Proc. Natl. Acad. Sci. USA 90: 5638-42; Kim & Little (1992)

15    15    Science 255: 203-6. Likewise, AraC hybrid homodimers should be unable to bind the hybrid operator because the low-affinity I2 half site does not bind AraC on its own. Reeder, T. & Schleif, R. (1993) J. Mol. Biol. 231, 205-218. Thus, activation would reflect heterodimerization of the hybrids through their zip domains, followed by cooperative binding of the heterodimer to

20    20    the LexAOp-I2 hybrid operator. This approach was unsuccessful despite many attempts to optimize DNA binding of the heterodimers (e.g. spacings of 4, 11 or 22 bp between the operator half sites). In some instances, a low level of activation of the lac gene by the AraC hybrid alone was in fact marginally repressed upon coexpression of an interacting LexA

25    25    hybrid, which indicated to us that the heterodimer subunits may actually interfere with each others binding to the closely spaced operator half sites. We also made various attempts at replacing zip-LexADNA and its site with a zip-RhaSDNA hybrid and RhaS operator half site, again without success. RhaS is a close structural and functional homologue of AraC; see Ramos et

30    30    al. (1990) Nucleic Acids Res. 18: 2149-52. The results suggest that an AraC monomer at I2 may not be sufficient for activation of the araBAD promoter.

**Repression of an activated lac reporter gene by heterodimeric E. coli AraC and LexA hybrid proteins in an E. coli two-hybrid system.**

As a new approach, we decided to set up a repression system where  
5 an initial activation event by AraC hybrids would be inhibited after an interacting LexA hybrid binds at LexAOp. As an AraC operator, we used the high affinity I1-I1 site which can bind dimeric and monomeric AraC molecules such as wild-type AraC and AraC<sub>DNA</sub>, respectively. (Figure 2A; Bustos & Schleif (1993), Proc. Natl. Acad. Sci. USA 90: 5638-42; Reeder, T. &  
10 Schleif, R. (1993) J. Mol. Biol. 231, 205-218. Since I1-I1 can also mediate activation of araBAD-lac by those dimeric and monomeric AraC molecules (Bustos & Schleif (1993), Proc. Natl. Acad. Sci. USA 90), we reasoned that this would allow us to examine hybrid interactions even if the fusion partner of AraC<sub>DNA</sub> fails to self-associate. Since the activation-based  
15 approach towards our E. coli two-hybrid system may have failed because the zip-AraC<sub>DNA</sub> and zip-LexA<sub>DNA</sub> heterodimer subunits may be unable to simultaneously bind closely spaced LexAOp and AraC operator sites (see above), we increased the spacing of the AraC and LexA operators by placing LexAOp several helical turns or at least 50 bp downstream of the promoter  
20 within the 5' untranslated leader of the lacZ reporter gene (Figure 2A). We hoped that zip-AraC<sub>DNA</sub> hybrids bound at the I1-I1 operator would mediate cooperative binding of zip-LexA<sub>DNA</sub> at LexAOp in an interaction-dependent manner. Operator binding of the LexA hybrid would presumably require looping of the DNA between the operators so that the operator sites are  
25 sufficiently close to each other for simultaneous binding of the heterodimer subunits. Finally, the approach assumes that twisting of the DNA between the operators would help to position the operators relative to each other for optimal access by the heterodimer subunits (Figure 3). Presumably, the operator-bound LexA hybrid interferes with activation by  
30 blocking the progression of RNAP (Elledge & Davis (1989), Genes & Dev. 3: 185-197), by excluding it from the promoter inside the DNA loop, (Figure 3; See Mueller et al. (1996), J. Mol. Biol. 257: 21-9; Choy & Adhya (1992),

Proc. Natl. Acad. Sci. USA 89: 11264-8; Huo *et al.* (1988), Proc. Natl. Acad. Sci. USA 85: 5444-8), or by sterically interfering with the AraC-RNAP contacts.

We inserted three tandem copies of LexAO<sub>p</sub> into either the XbaI or 5 HindIII sites in the 5' untranslated leader of the lacZ gene on the reporter plasmid (Figure 2A). We also inserted the lambda H' site (IHFO<sub>p</sub>; Yang & Nash (1994), Proc. Natl. Acad. Sci. 91: 12183-12187) for the E. coli IHF protein into the SacI site between I1-I11 and LexAO<sub>p</sub> (Figs. 2A and 3). Since 10 IHF enhances DNA looping by bending DNA (Nash (1990) Trends Biochem. Sci. 15: 222-7. Pratt *et al.* (1996) Mol. Microbiol. 20: 911-7; Hoover *et al.* (1990) Cell 63: 11-22), we hoped that IHF-mediated DNA looping in our system would enhance repression.

**Homo- and heterodimeric protein interactions can mediate repression.**

We have tested our general scheme with both a zip-AraCDNA/zip- 15 LexADNA hybrid pair, and a second set of hybrids in which the interaction domains of Fos and Jun were fused to AraCDNA and LexADNA, respectively (Figure 2B). Unlike the c/EBP zip domain, Fos does not self-associate. Halazonetis *et al.* (1988) Cell 55: 917-24. Repression with the 20 hybrid pair would provide direct evidence that the oligomerization state of AraC hybrids is not important for repression or the initial activation event at I1-I11. The interaction domain of Jun can mediate dimerization when fused to LexADNA and the Fos domain can disrupt Jun-LexADNA dimers *in vitro*. Doerr *et al.* (1991) Biochem. 30: 9657-64.

We are able to most clearly demonstrate repression in our LexA- 25 AraC system when we use the zip-LexADNA/zip-AraCDNA hybrid pair with a reporter plasmid that contains IHFO<sub>p</sub> inserted at SacI and LexAO<sub>p</sub> at XbaI (P3IHFLexAO<sub>p</sub>X). In Figure 4A, we note that this reporter plasmid expresses 160 units of β-galactosidase when present alone. This basal level of expression is substantially above the approximately 40 units of activity 30 noted with the reporter plasmid in the absence of LexOp. See table 3 below and Reeder and Schleif (1993), J. Mol. Biol. 231: 205-218. We believe that the enhanced expression represents a contribution made by the fortuitous

promoter(s) introduced with the LexOp sites (see below). The basal level of expression is elevated to 620 units in the presence of zip-AraCDNA. In the presence of both zip-LexADNA and zip-AraCDNA, expression drops to 220 units demonstrating repression. If we subtract the basal level of expression  
5 (160 units) from that noted with zip-AraCDNA (620 units), we obtain a value of 460 units which we believe represents the expression resulting from the activation of the AraC-controlled promoter. Likewise, when the basal level is subtracted from the expression seen when zip-LexADNA and zip-AraCDNA are coexpressed, we obtain a value which represents the  
10 reduced residual activation noted with the interacting pair (60 units). We define the ratio of these two values as the repression factor;  $460/60 = 7.5$  in this case. In Figure 4B, we show that repression is also seen with another interacting fusion pair, Jun-LexADNA and Fos-AraCDNA. In this instance a repression factor of 4.2 is noted. We will demonstrate below that  
15 optimal repression requires interacting protein domains in the chimera, appropriately positioned DNA binding sites, and IHF with its appropriate binding site.

**IHF can enhance repression.**

We can show that IHF is involved in repression by noting reduced  
20 repression when IHF is removed by mutation or by deletion of the IHF binding site. In Table 1 we see that repression with both the Fos-AraCDNA/Jun-LexADNA and zip-AraCDNA/zip-LexADNA hybrid pairs is reduced by a factor of 2 with a reporter construct that contains LexAOp at the XbaI site, but lack IHFOp at the SacI site (P3LexAOpX vs  
25 P3IHFLexAOpX). Surprisingly, when LexOp site is placed at the HindIII site, the presence of IHFOp site is without an effect (Table 1; P3LexAOpH vs P3IHFLexAOpH). Apparently the ability of IHF to facilitate repression is dependent on the detailed geometry of our putative repression loop; its aids in repression with the small 117 bp loop when LexAOp is at XbaI, but  
30 is without effect with the larger 180 bp loop that is formed with LexAOp at HindIII.

The IHFOP oligonucleotide insert has a length of 40 bp. This corresponds to roughly 3.6 turns of the helix (Figure 2A; Lee & Schleif (1989), *Proc. Natl. Acad. Sci.* 86: 476-480), which raised the possibility that IHFOP enhanced repression with P3LexAOpX because of altered relative positioning of the interacting hybrids on the face of the helix. When we 5 mutated the bases comprising the consensus IHF binding site while maintaining the length of IHFOP (Materials & Methods), repression with both hybrid pairs is reduced to values similar to those seen without the IHFOP insert. Furthermore, we are able to show that repression with the 10 Fos-AraCDNA/Jun-LexADNA and P3IHFLexAOpX is completely eliminated in a strain carrying the chromosomal himD::cat mutation which inactivates the HimD subunit of IHF (compare Figs. 4B and 4C).

Based on these results, optimal repression with P3LexAOpX therefore occurs when an IHFOP site is present at SacI and the strain is 15 himD<sup>+</sup>. Removal of IHFOP reduces repression with zip-AraCDNA/zip-LexADNA and this same manipulation or removal of HimD by mutation completely eliminate repression with Fos-AraCDNA/Jun-LexADNA (see below for details). We believe that the enhancement of repression by IHF is the result of DNA loop formation which aids in bringing the AraC and 20 LexA operator sites into spatial proximity for enhanced binding by heterodimeric hybrids.

**Repression requires hybrid interaction.**

Full repression is only noted when the fused domains of the chimeric protein partners are able to interact, as in our examples with the 25 zip and Jun/Fos pairs (Figure 4A and Table I). When noninteracting pairs are present, repression is not detectable (e.g. when wild type AraC or AraCDNA are paired with zip-LexADNA or when AraCDNA is paired with Jun-LexADNA; Figure 4D and Table 2). Repression is also absent when normally interacting protein partners (Jun/Fos) have their interaction 30 disrupted by the introduction of a specific mutation, as in our example with the Fos IVM54 point mutation (Figure 4E) which has been previously shown to block Fos/Jun heterodimerization. Ryseck *et al.* (1990), *Oncogene*

5: 1091-3. Western blot analysis has not given any indications that the differences in repression levels noted with the various reporter construct and hybrid pairs can be contributed to altered hybrid protein levels; the repression described above therefore does not simply reflect altered  
5 relative amounts of the AraC and LexA proteins.

**Reducing background Lac reporter activity improves repression.**

Table 3A shows that the reporter plasmids of this study produce close to 10-fold higher background (AraC-independent) lac transcription than the parent plasmid P3. Inspection of the LexAOp sequence revealed  
10 that each LexA half site contains a putative -10 E. coli promoter sequence (Figure 2; 5'-TACAGT-3') that might act with upstream -35 promoter-like sequences to promote AraC independent lac transcription. We changed the putative -10 sequences of P3IHFLexAOpH to 5'-ACCAGT-3' to eliminate -10 activity while leaving intact the CAG trinucleotide that constitutes the  
15 core consensus LexA binding site. Lewis *et al.* (1994), *J. Mol. Biol.* 241: 507-523. The manipulation decreased background Lac activity by about 20-fold (290 Miller units of Lac activity are reduced to about 17 Miller units). To demonstrate that the new reporter plasmid (P3IHFLexOpmH) permits increased repression, we assayed repression with the zip-AraCDNA / zip-LexADNA interacting pair. We observe improved repression: the 5-fold repression with P3IHFLexAOpH (Table 1) increases to about 50-fold with P3IHFLexAOpmH (Table 3B and Fig 4). Repression remained negligible with the non-interacting AraCDNA / Jun-LexADNA, AraCDNA / zip-LexADNA and AraC / zip-LexADNA hybrid pairs (Table 3B).

25 Zip-AraCDNA and wild-type AraC generate about 1000 Miller units of Lac activity with P3IHFLexAOpmH plasmid. This 60-fold activation of lac above background compares favourably with the approximately 100-fold activation reported for the parent plasmid P3. Reeder & Schleif (1993), *J. Mol. Biol.* 231: 205-218. Therefore, the low activation seen with the  
30 unmodified reporter plasmids reflects interference by the high background of AraC-independent Lac activity associated with these plasmids. As

demonstrated with P3IHFLexAOpmH, this problem can be eliminated by reducing AraC-independent background lac transcription.

**An improved two-hybrid system.**

The affinity for LexA of the nonconsensus LexA site in LexAOpm is  
5 unknown, but it is predicted to be less than the affinity of LexAOp. Lewis *et*  
*al.* (1994), *J. Mol. Biol.* 241: 507-523. LexAOpm improves repression with  
zip-AraC<sub>DNA</sub>/zip-LexA<sub>DNA</sub> (see above), but almost completely abolishes  
repression by Fos-AraC<sub>DNA</sub>/Jun-LexA<sub>DNA</sub> (not shown). These results most  
likely reflect the fact that foreign fusion domains can alter the operator  
10 affinity of LexA<sub>DNA</sub> (Golemis & Brent (1992) *Mol. Cell. Biol.* 12: 3006-3014)  
and they suggest that Jun-LexA<sub>DNA</sub> has a lower affinity for LexA operators  
than zip-LexA<sub>DNA</sub>.

These observations suggest that repression in our system can  
generally be maximized by improving the affinity of LexA operators.  
15 Reversal of the AraC and LexA fusion partners or fusions to the C-  
terminus of wild-type LexA may also minimize inhibitory effects of the  
fusion domain on LexA<sub>DNA</sub>. Golemis & Brent (1992) *Mol. Cell. Biol.* 12:  
3006-14. High affinity LexA operators with low endogenous promoter  
activity should minimize interference of background lac activity with the  
20 repression signal. Repression, may also be enhanced by overexpressing  
LexA hybrids. Finally, AraC<sub>DNA</sub> activates less efficiently than full-length  
AraC. Bustos & Schleif (1993) *Proc. Natl. Acad. Sci. USA* 90: 5638-42.  
Hybrids of full-length AraC may therefore improve repression by  
providing a greater activation signal.

25 To test if manipulation of these parameters improves repression  
with the Fos/Jun interaction pair, we generated reporter plasmid  
P3IHFLexAOimpX by replacing LexAOpmX with three tandem direct  
repeat impAB LexA operator half sites which are known to have a high  
affinity for LexA (Lewis *et al.* (1994), *J. Mol. Biol.* 241: 507-523) and a low  
30 endogenous promoter activity of approximately 10 Miller units. We also  
fused Jun to full-length AraC and Fos to the C-terminus of full-length  
LexA expressed from the strongly anhydrotetracycline-inducible tet

promoter (see Materials & Methods) to generate the Jun-AraC/Fos-LexA hybrid pair in Figure 2B (see Materials & Methods). We observe a repression ratio of about 11-fold. This constitutes a three-fold improved level of repression when compared with our results for the Fos-  
5 AraC<sub>DNA</sub>/Jun-LexA<sub>DNA</sub> hybrid pair (compare Table 1 and Figure 5A). No significant repression is noted with a control hybrid pair (AraC/Fos-LexA; Figure 5B). Figure 5A also shows that repression decreases as expression of fos-lexA from the tet promoter is reduced. The results suggest that our strategies can successfully improve repression with any pair of interaction  
10 domains. The dependence of repression on the relative amounts of the AraC and LexA hybrids also provides a third line of evidence that repression reflects heterodimerization.

#### Interaction cloning

Our system detects the interaction between domains of the HPV E6  
15 and human E6AP proteins that interact in the yeast two-hybrid system.  
Huibregtse et al. (1991), EMBO J. 10: 4129-35. With 0.5 mM IPTG to induce E6AP-AraC<sub>DNA</sub>, and 50ng/ml of anhydrotetracycline to induce E6-LexA, we obtain a repression ratio of at least 30-fold. Repression is essentially absent with the non-interacting control pair AraC<sub>DNA</sub>/E6-LexA or when the LexA  
20 operator is deleted from the P3IHFLexAimpX reporter plasmid (repression ratios of about 1.5 in each case). On solid medium with the LacZ indicator X-gal, repression with E6AP-AraC<sub>DNA</sub>/E6-LexA generates pale blue colonies whereas dark blue colonies result with the controls. To evaluate if this signal differential can be exploited for interaction cloning, we mixed a  
25 BamH1 E6 fragment with an approximately 100-fold molar excess of total BamH1 digested E. coli chromosomal DNA. The mixture was ligated into the BamH1 site at the 3' end of lexA in our LexA expression vector where the correctly oriented E6 DNA is predicted to be in frame with lexA. After transformation into the reporter strain containing E6AP-AraC<sub>DNA</sub>  
30 construct, more than 95% of the transformants were blue on media with X-gal, 0.5 mM IPTG and 50ng/ml AT. The remaining white colonies presumably contained the LexA-E6 fusion or they resulted from loss of the

reporter plasmid which appears to occur at a low frequency. To avoid picking colonies that have lost the reporter gene and to try to specifically isolate colonies with LexA-E6 fusions, we picked 12 white colonies that eventually slowly turned blue. To further enrich for colonies containing 5 LexA-E6 hybrids, plasmid DNA from all 12 colonies was pooled, retransformed into the reporter strain and plated exactly as above. The majority of transformants were now pale blue and PCR mapping of 16 randomly isolated pale blue transformants showed that all contained the E6AP coding fragment in the correct orientation behind lexA. The results 10 show that a simple protocol can be used to efficiently isolate specific interactors of the AraC fusion partner by interaction cloning when only approximately 1 in 200 clones is expected to contain the E6-fragment fused to lexA in the correct orientation.

#### DISCUSSION

15 The E. coli two-hybrid system described here signals heterodimerization of AraC and LexA hybrids through repression of a lacZ reporter gene. Repression clearly requires two-hybrid interaction because noncognate hybrid pairs do not generate repression and because interaction cloning with E6AP-AraC as a bait could be used to specifically 20 isolate E6 from an E6-spiked DNA library fused behind LexA. Repression also requires the LexA site and can be further enhanced by IHF between the AraC and LexA operators of the reporter plasmid, suggesting the involvement of DNA looping between operator-bound and heterodimeric hybrids (Figure 3). Large repression loops should be less dependent on IHF 25 because large loops bend and twist more easily (Rippe *et al.* (1995), Trends in Biochem. Sci. 20: 500-506.

35; Wang & Giaever (1988), Science 240: 300-306) and we have shown that IHF enhances repression with P3LexAOpX (small loop size of about 77 bp), but not with P3LexAOpH (large loop size of about 140 bp) (Table I). IHF- 30 mediated DNA looping normally modulates promoter activity at other bacterial loci (Nash (1990) Trends Biochem. Sci. 15: 222-7; Pratt *et al.* (1996) Mol. Microbiol. 20: 911-7; Hoover *et al.* (1990) Cell 63: 11-22). The

involvement of IHF in these distinct systems suggests that it functions as an independent module, thus providing a rationale for its effect in the artificial context of the E. coli two-hybrid system.

Repression may reflect exclusion of RNAP from the promoter  
5 inside the loop (Mueller *et al.* (1996), *J. Mol. Biol.* 257: 21-9; Choy & Adhya (1992), *Proc. Natl. Acad. Sci. USA* 89: 11264-8; Huo *et al.* (1988), *Proc. Natl. Acad. Sci. USA* 85: 5444-8). A second mechanism of repression is suggested by the two-fold repression with zip-AraC<sub>DNA</sub>/zip-LexA<sub>DNA</sub> and the P3 reporter plasmid (Table I). Since dimeric AraC molecules activate araBAD  
10 more efficiently from I1-I1 than monomeric AraC molecules (Reeder, T. & Schleif, R. (1993) *J. Mol. Biol.* 231: 205-218), zip-LexA<sub>DNA</sub> appears to cause repression with P3 by disrupting dimeric zip-AraC<sub>DNA</sub> through heterodimerization (Figure 6). Consistent with this interpretation, Fos-AraC<sub>DNA</sub> is monomeric and Jun-LexA<sub>DNA</sub> fails to repress with plasmid P3  
15 (Table I). Hybrid interaction-dependent DNA looping may also cause repression by altering DNA bending of the araBAD promoter region (Nickerson & Achberger (1995), *J. Bacteriol.* 177: 5756-61), or it may reflect physical interference of the repression loop (Chatterjee *et al.* (1997), *Proc. Natl. Acad. Sci. USA* 94: 2957-62) or the operator-bound LexA hybrid  
20 (Elledge & Davis (1989), *Genes & Dev.* 3: 185-197) with the activation functions of DNA-bound AraC or RNAP. Indeed, the geometry of AraC-RNAP interactions is likely to be complex because of the involvement of the RNAP  $\alpha$ -subunit (Giffard & Booth (1988), *Mol. Gen. Genet.* 214: 148-152). The involvement of AraC- $\alpha$  interactions may explain why we were  
25 unable to achieve activation with our AraC/LexA hybrid heterodimers.  
**General utility of the E. coli two-hybrid system.**

All AraC<sub>DNA</sub> hybrids in this report retain the ability to activate the araBAD promoter. We have also noted activation with larger fusion partners from the human papilloma virus E2 protein (Mueller & Sapp (1996), *Virology* 219: 247-256) and the eucaryotic DP1 (Girling *et al.* (1993), *Nature* 352: 83-87) and E47 (Kadesch (1993), *Cell Growth & Different.* 4: 49-55) proteins. Some fusion partners abolish the ability of AraC<sub>DNA</sub> to

activate, but we generally observe good activation with fusions to the N-termini of wild-type AraC or zip-AraC<sub>DNA</sub>. These findings are consistent with the organization of AraC into activation and dimerization domains (Eustance *et al.* (1994), *J. Mol. Biol.* 242: 330-338) and they suggest that AraC 5 hybrids with an intact activation function can be obtained with most fusion partners. The suitability of LexA for generating hybrid proteins with good LexA operator binding ability is well documented (*e. g.* Allen *et al.* (1995), *Trends in Biochem. Sci.* 20: 511-16).

Some eucaryotic proteins fail to properly fold in *E. coli* (Cleland 10 (1993) in *Protein Folding: in Vivo and in Vitro*, ed. J. Cleland (ACS), pp. 1-21) and this represents a potential limitation. We believe the system will prove to be particularly useful with procaryotic proteins, eucaryotic proteins that can be obtained in active form from *E. coli*, and smaller defined domains of eucaryotic proteins such as the E6/E6AP protein pair.

15 Mechanistically, the system differs fundamentally from the yeast two-hybrid system where protein interaction is indicated by reporter gene activation, but like the yeast system, it can detect homo- and heterodimeric interactions as demonstrated with the zip/zip and Fos/Jun protein pairs. These features distinguish these two systems from other *E. coli* systems 20 that principally detect homodimeric interactions (Doerr *et al.* (1991) *Biochem.* 30, 9657-9664; Marchetti *et al.* (1995) *J. Mol. Biol.* 248, 541-550; Jappelli & Brenner (1996) *J. Mol. Biol.* 259, 575-578). The DNA loop between the AraC and LexA operators may be a key factor that contributes 25 to the effectiveness of our two-hybrid system. Theoretical considerations (Rippe *et al.* (1995), *Trends in Biochem. Sci.* 20: 500-506; Wang & Giaever (1988), *Science* 240: 300-306) suggest that twisting and bending of a DNA loop reduces steric constraints that may prevent heterodimer subunits from binding closely adjacent half sites. This flexibility is a feature that is missing in the *E. coli* systems that detect protein homodimerization.

30 The potential for large interaction-dependent variations in reporter gene activity points to the basic usefulness of the system for the same applications for which the yeast two-hybrid system has been extensively

used. These applications include the mapping of protein interaction domains and 'interaction cloning' and we have shown here that the current system version is probably adequate for isolating interactors from small insert libraries prepared from bacterial or yeast genomes.

- 5 Replacement of the lac reporter gene with a gene whose expression is toxic (Wall (1996), BioTechniques 20: 690-693), or the use of a toxic substrate for  $\beta$ -galactosidase (Davis & Jacob (1968) J. Mol. Biol. 36: 413-27), should permit positive selection of interacting clones from large mammalian insert libraries fused to LexA. Another application, the high-throughput
- 10 screening for therapeutically active small molecule modulators of protein interactions (Gibbs & Oliff (1994) Cell 79: 193-198), may benefit from the faster growth rate and increased permeability of mutant E. coli strains (Higgins (1993) Curr. Opin. in Cell Biol. 5: 684-687; Nakamura & Suganuma (1972) J. Bacteriol. 110: 329-35). Furthermore, E. coli does not
- 15 require nuclear localization of hybrid proteins which has been problematic with non-nuclear target proteins in yeast (reviewed in Allen et al. (1995), Trends in Biochem. Sci. 20: 511-16). We believe that the system described here is a useful alternative to another E. coli system, which is mechanistically entirely different from the system described here, and for
- 20 which applications similar to those of the yeast two-hybrid system have been proposed (Dove et al. (1997), Nature 386: 627-30).

**Table 1**

Summary of repression ratios with hybrid pairs containing cognate interaction domains. See text for calculation of ratios. 'AraC' and 'LexA' refer to the DNA binding domains of these proteins.

5

**Reporter Construct**

	P3	P3 LexAOpH	P3IHF LexAOpH	P3 LexAOpX	P3IHF LexAOpX
zip-AraC	2.1	4.9	4.1	2.2	7.5
zip-LexA					
Fos-AraC	1.1	2.0	1.9	2.3	4.2
Jun-AraC					

**Table 2**

Summary of repression ratios with hybrid pairs predicted to be unable to interact. 'AraC' and 'LexA' are defined in Table 1. AraC-wt = wild-type AraC. ND = not determined.

**Reporter Construct**

	P3	P3 LexAOpH	P3IHF LexAOpH	P3 LexAOpX	P3IHF LexAOpX
AraC-wt	1.2	1.1	1.0	1.2	1.2
zip-LexA					
AraC	1.2	1.0	1.1	1.1	1.2
zip-LexA					
AraC	ND	ND	ND	ND	1.1
Jun-LexA					

**Table 3**

Enhanced repression with reduced background reporter gene activity.

(A) Summary of background Miller units of Lac activities generated by plasmid P3 and the reporter plasmids of Table 1.

5 (B) Repression by zip-AraC<sub>DNA</sub>/zip-LexA<sub>DNA</sub> and control hybrid pairs with the P3IHFLexAOpmH reporter plasmid. AraC, LexA and AraC(wt) are as in Figure 4.

**Reporter Construct**

<b>A</b>	P3	P3 LexAOpH	P3IHF LexAOpH	P3 LexAOpX	P3IHF LexAOpX
Miller Units	36	300	290	270	160

10

**Hybrid Pair**

<b>B</b>	zip-AraC + zip-LexA	AraC + zip-LexA	AraC(wt) + zip-LexA	AraC + Jun-LexA
repr. ratio with P3IHF- LexAOpmH	53	0.9	1.06	1.09

The abbreviations used in this specification are defined as follows.

AT	anhydrotetracycline
bp	base pairs
CPRG	Chlorophenolred- $\beta$ -D galactopyranoside
5	cDNA complementary DNA
DNA	deoxyribonucleic acid
IPTG	isopropylthiogalactoside
kb, kbp	kilobase pairs
MU	Miller units
10	RNA ribonucleic acid
wt	wild type

What is claimed is:

1. A prokaryotic host cell comprising:
  - (a) a fusion protein having
    - (i) a first DNA-binding domain and
    - (ii) a first protein-interacting domain;
  - (b) a fusion protein having
    - (i) a second DNA-binding domain and
    - (ii) a second protein-interacting domain capable of binding to the first interacting domain; and
  - (c) a nucleic acid molecule having a reporter gene operatively linked to
    - (i) a promoter,
    - (ii) a first operator site capable of binding to the first DNA-binding domain, located upstream of the promoter, and
    - (iii) a second operator site capable of binding the second DNA-binding domain, located downstream of the promoter and the first operator site;
- 20 wherein binding of the first interacting domain to the second interacting domain is signaled by altered expression of the reporter gene.
2. A nucleic acid having a having a reporter gene operatively linked to
  - (a) an AraC operator site,
  - (b) a promoter, and
  - (c) a LexA operator site.
3. A reporter vector comprising the nucleic acid of Claim 2.

30

4. A process for detecting inhibition or enhancement of binding of a first interacting domain with a second interacting domain, which comprises:

- 5 (a) treating a culture of cells of Claim 1 with a test substance, and  
(b) screening for altered expression of the reporter gene.

5. A process for detecting a cell comprising a test domain that interacts with a known domain, which comprises:

- 10 (1) generating a library of cells in which each cell comprises:  
15 (a) a fusion protein having  
      (i) a first DNA-binding domain and  
      (ii) a known domain;  
      (b) a fusion protein having  
          (i) a second DNA-binding domain and  
          (ii) a test domain; and  
      (c) a nucleic acid molecule having a reporter gene  
          operatively linked to  
          (i) a promoter,  
20          (ii) a first operator site capable of binding to the first  
          DNA-binding domain, located upstream of the  
          promoter, and  
          (iii) a second operator site capable of binding the  
          second DNA-binding domain, located  
          downstream of the promoter of the reporter gene;  
25 (2) detecting cells exhibiting altered expression of the reporter  
      gene, which signals interaction between the known domain  
      and the test domain.

6. The cell of Claim 1, wherein:
  - (a) the first DNA-binding domain is an AraC DNA binding domain,
  - (b) the first operator site is an AraC operator site,
  - 5 (c) the second DNA-binding domain is a LexA DNA binding domain, and
  - (d) the second operator site is a LexA operator site.
7. The process of Claim 4, wherein:
  - (a) the first DNA-binding domain is an AraC DNA binding domain,
  - (b) the first operator site is an AraC operator site,
  - 10 (c) the second DNA-binding domain is a LexA DNA binding domain, and
  - (d) the second operator site is a LexA operator site.
8. The process of Claim 5, wherein:
  - (a) the first DNA-binding domain is an AraC DNA binding domain,
  - 20 (b) the first operator site is an AraC operator site,
  - (c) the second DNA-binding domain is a LexA DNA binding domain, and
  - (d) the second operator site is a LexA operator site.
- 25 9. The cell of Claim 1, wherein the nucleic acid molecule further comprises more than one second operator site operatively linked to the reporter gene.
10. The process of Claim 4, wherein the nucleic acid molecule  
30 further comprises more than one second operator site operatively linked to the reporter gene.

11. The process of Claim 5, wherein the nucleic acid molecule further comprises more than one second operator site operatively linked to the reporter gene.
- 5 12. The nucleic acid molecule of Claim 2, further comprising more than one LexA operator half site operatively linked to the reporter gene.
13. A reporter vector comprising the nucleic acid of Claim 12.
- 10 14. The cell of Claim 1, wherein the reporter gene is selected from nucleic acids encoding  $\beta$ -galactosidase, antibiotic resistance genes, and toxic genes.
- 15 15. The nucleic acid of Claim 2, wherein the reporter gene is selected from nucleic acids encoding  $\beta$ -galactosidase, antibiotic resistance genes, and toxic genes.
16. A reporter vector comprising the nucleic acid of Claim 15.
- 20 17. The process of Claim 4, wherein the reporter gene is selected from nucleic acids encoding  $\beta$ -galactosidase, antibiotic resistance genes, and toxic genes.
- 25 18. The process of Claim 5, wherein the reporter gene is selected from nucleic acids encoding  $\beta$ -galactosidase, antibiotic resistance genes, and toxic genes.
19. The cell of Claim 1, wherein the cell is a bacterial cell.
- 30 20. The process of Claim 4, wherein the cell is a bacterial cell.
21. The process of Claim 5, wherein the cell is a bacterial cell.

5

22. The cell of Claim 1, wherein the cell is an E. coli cell.
23. The process of Claim 4, wherein the cell is an E. coli cell.
24. The process of Claim 5, wherein the cell is an E. coli cell.

1 / 7

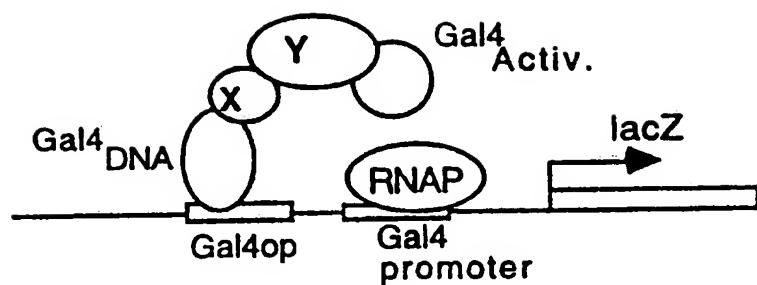


FIG. 1A

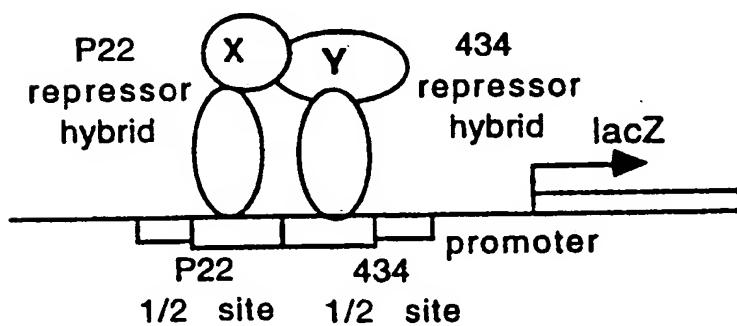
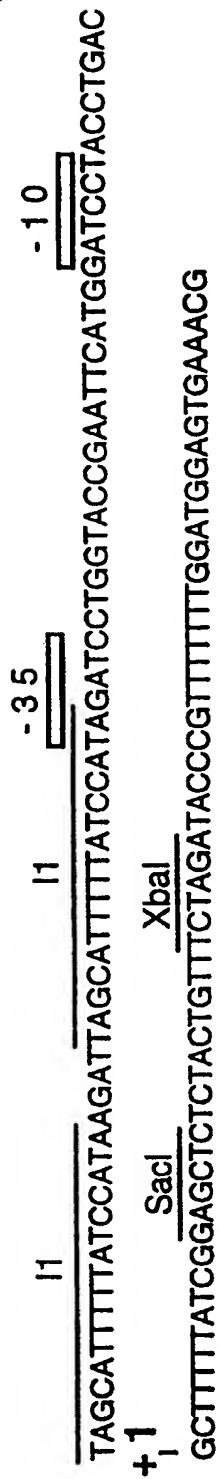


FIG. 1B

2 / 7



LexAOp for XbaI:  
 XbaI-SacI **ATATACAGTA ATATACAGTA ATATACAGTA C-Saci-XbaI**

LexAOp for HindIII:  
 HindIII-T-XbaI-SacI **ATATACAGTA ATATACAGTA ATATACAGTA C-Saci-XbaI-HindIII**

Lambda H' IHF site for SacI:  
 SacI-ATAAAAAGCATTGCTTATCAATTGGCAAGG-Saci

**FIG. 2A**

3 / 7

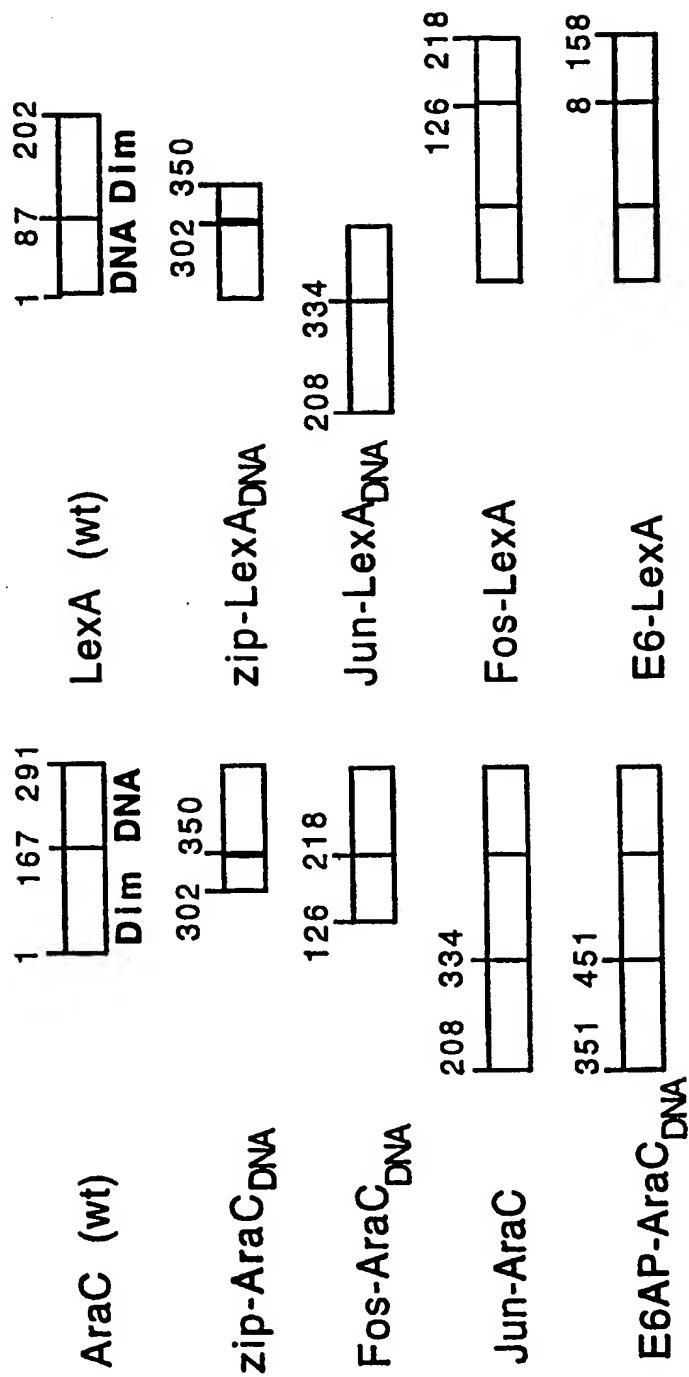


FIG. 2B

4 / 7

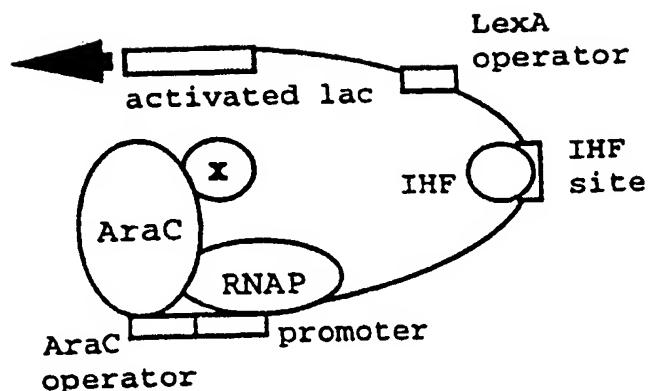


FIG. 3A

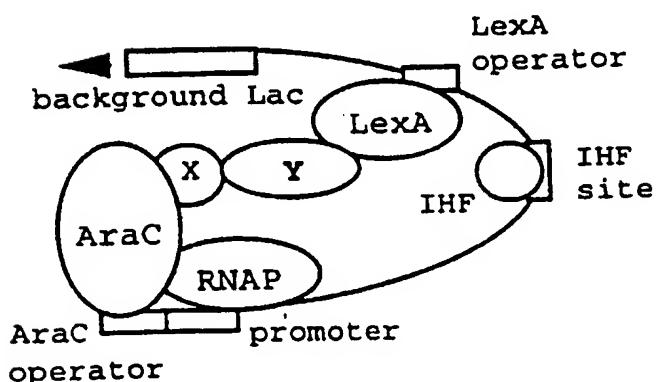


FIG. 3B

5 / 7

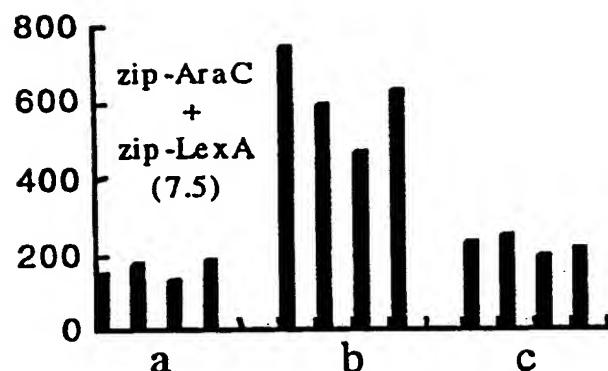


FIG. 4A

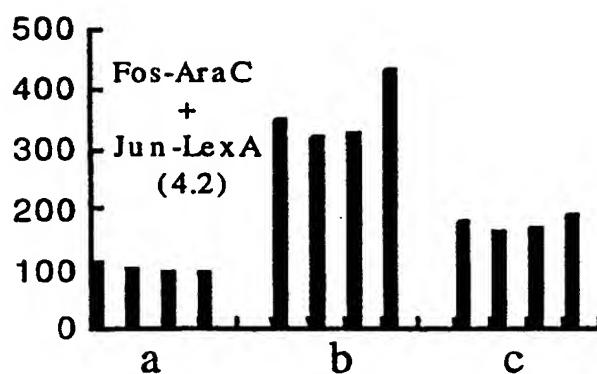


FIG. 4B

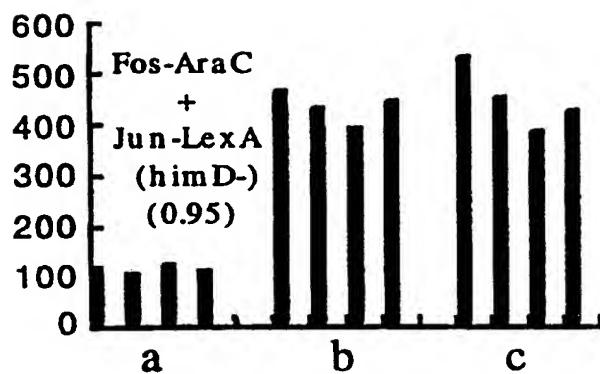


FIG. 4C

6 / 7

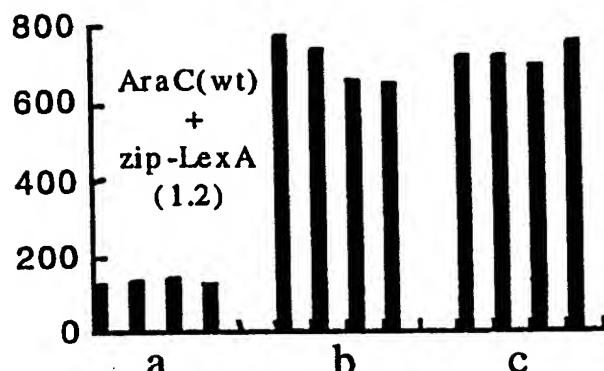


FIG. 4D

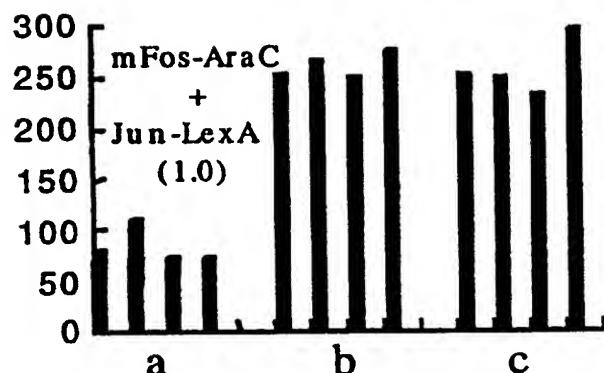


FIG. 4E

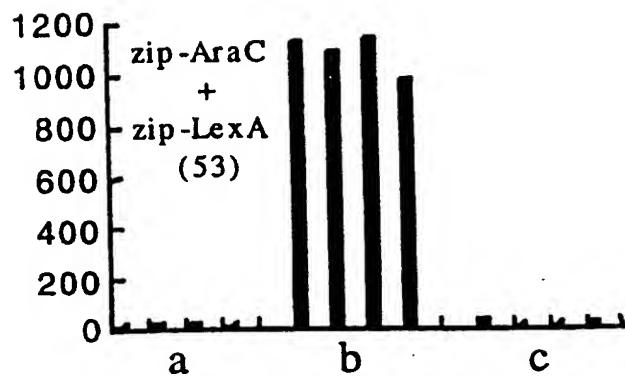


FIG. 4F

7 / 7

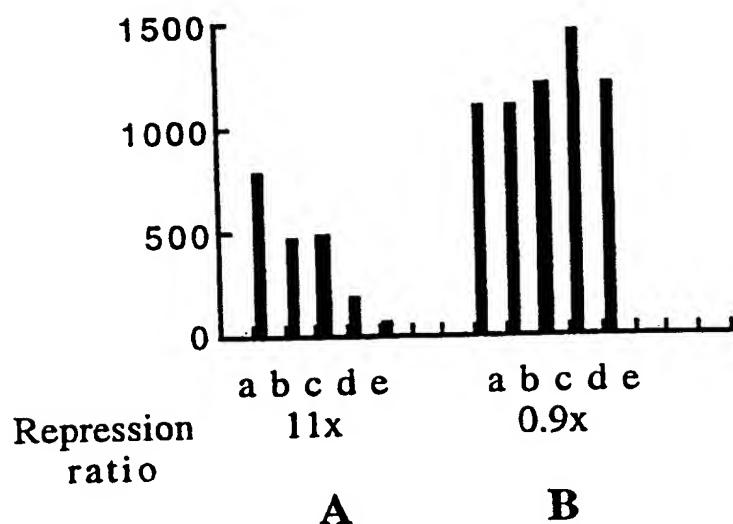


FIG. 5

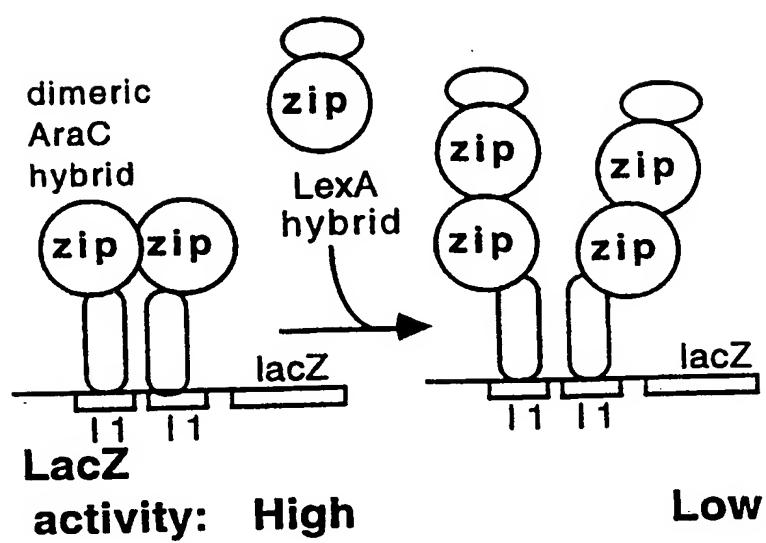


FIG. 6

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US97/22703

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :C07H 21/04; C12Q 1/68; C12N 1/21, 15/70; G01N 33/53  
US CL :435/6, 7.1, 252.3, 252.33, 320.1; 536 23.1

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 7.1, 252.3, 252.33, 320.1; 536 23.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)  
Please See Extra Sheet.

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	LUBAN et al. The yeast two-hybrid system for studying protein-protein interactions. Curr. Opin. Biotechnol. 1995, Vol. 6, pages 59-64, see entire document.	1-24
A	FIELDS et al. The two-hybrid system: an assay for protein-protein interactions. Trends Genet. August 1994, Vol. 10, No. 8, pages 286-292, see entire document.	1-24

Further documents are listed in the continuation of Box C.

See patent family annex.

Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"A"	document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means		
"P" document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search

27 FEBRUARY 1998

Date of mailing of the international search report

08 APR 1998

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231  
Facsimile No. (703) 305-3230

Authorized officer  
ROBERT SCHWARTZMAN  
Telephone No. (703) 308-0196

**INTERNATIONAL SEARCH REPORT**

International application No.  
PCT/US97/22703

**B. FIELDS SEARCHED**

Electronic data bases consulted (Name of data base and where practicable terms used):

STN: Medline, Biosis, CAPIus, Embase, WPIDS, JAPIO, PATOSEP , PATOSWO

APS

Search terms: two hybrid, DNA binding, procaryote, bacteria, lexA, araC

THIS PAGE BLANK (USPTO)

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**